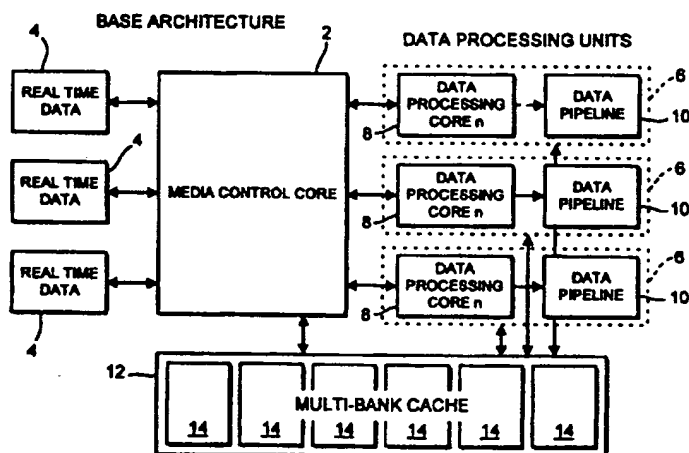




INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

| | | |
|---|----|---|
| (51) International Patent Classification ⁶ : G06F 15/80 | A1 | (11) International Publication Number: WO 97/38372 (43) International Publication Date: 16 October 1997 (16.10.97) |
| (21) International Application Number: PCT/GB97/00972 (22) International Filing Date: 4 April 1997 (04.04.97) (30) Priority Data: 9607153.5 4 April 1996 (04.04.96) GB (71) Applicant: VIDEOLOGIC LIMITED [GB/GB]; Home Park Estate, Kings Langley, Hertfordshire WD4 8LX (GB). (72) Inventors: WHITTAKER, James, Robert; 9B Kings Road, Berkhamsted, Hertfordshire HP4 3BD (GB). ROWLAND, Paul; 158B London Road, St. Albans, Hertfordshire AL1 1PQ (GB). (74) Agent: ROBSON, Aidan, John; Reddie & Grose, 16 Theobalds Road, London WC1X 8PL (GB). | | (81) Designated States: JP, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE). Published <i>With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i> |

(54) Title: A DATA PROCESSING MANAGEMENT SYSTEM



(57) Abstract

A data processing management system comprises one or more data inputs (34) and one or more data outputs (42). It also includes one or more data processing units (50), a data storage means (12) and a control core (2). The control means includes means for routing data between the data input, the data output, the data processing means and the data storage means in one or more programmable routing operations. It is able to cause the data processing means to commence a predetermined data processing operation. It is also able to repeatedly determine which routing operations and which data processing operations are capable of being performed and is then able to commence execution of at least one of the thus determined operations capable of being performed.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

| | | | | | | | |
|----|--------------------------|----|--|----|--|----|--------------------------|
| AL | Albania | ES | Spain | LS | Lesotho | SI | Slovenia |
| AM | Armenia | FI | Finland | LT | Lithuania | SK | Slovakia |
| AT | Austria | FR | France | LU | Luxembourg | SN | Senegal |
| AU | Australia | GA | Gabon | LV | Latvia | SZ | Swaziland |
| AZ | Azerbaijan | GB | United Kingdom | MC | Monaco | TD | Chad |
| BA | Bosnia and Herzegovina | GE | Georgia | MD | Republic of Moldova | TG | Togo |
| BB | Barbados | GH | Ghana | MG | Madagascar | TJ | Tajikistan |
| BE | Belgium | GN | Guinea | MK | The former Yugoslav Republic of Macedonia | TM | Turkmenistan |
| BF | Burkina Faso | GR | Greece | | | TR | Turkey |
| BG | Bulgaria | HU | Hungary | ML | Mali | TT | Trinidad and Tobago |
| BJ | Benin | IE | Ireland | MN | Mongolia | UA | Ukraine |
| BR | Brazil | IL | Israel | MR | Mauritania | UG | Uganda |
| BY | Belarus | IS | Iceland | MW | Malawi | US | United States of America |
| CA | Canada | IT | Italy | MX | Mexico | UZ | Uzbekistan |
| CF | Central African Republic | JP | Japan | NE | Niger | VN | Viet Nam |
| CG | Congo | KE | Kenya | NL | Netherlands | VU | Yugoslavia |
| CH | Switzerland | KG | Kyrgyzstan | NO | Norway | ZV | Zimbabwe |
| CI | Côte d'Ivoire | KP | Democratic People's Republic of Korea | NZ | New Zealand | | |
| CM | Cameroon | KR | Republic of Korea | PL | Poland | | |
| CN | China | KZ | Kazakhstan | PT | Portugal | | |
| CU | Cuba | LC | Saint Lucia | RO | Romania | | |
| CZ | Czech Republic | LI | Liechtenstein | RU | Russian Federation | | |
| DE | Germany | LK | Sri Lanka | SD | Sudan | | |
| DK | Denmark | LR | Liberia | SE | Sweden | | |
| EE | Estonia | | | SG | Singapore | | |

- 1 -

A DATA PROCESSING MANAGEMENT SYSTEM

This invention relates to a data processing management system of the type which can be used with real time multimedia inputs and processing.

5 BACKGROUND TO THE INVENTION

The user interface to computers has continually evolved from teletypes to keyboard and character terminals to the (graphical user interface) GUI which is currently the standard interface for the majority
10 of computer users. This evolution is continuing with sound and 3D graphics increasingly common and 3D sound and virtual reality emerging. It's common thread is an increase in the complexity of the human computer interface achieved by an accompanying
15 increase in the types of data presented to the user. (personal computer) PC applications are taking advantage of this shift and are increasingly relying on the availability of sound and 3D graphics in order to achieve their full potential.

20 This has resulted in chip and board suppliers offering products with combined functionality designed to handle more than one data type e.g. 2D graphics and sound or 2D and (motion picture experts group) MPEG playback. It is important to note that
25 these products to date use separate functional units for each data type.

More recently, programmable SIMD (Single Instruction Multiple Data) architectures (e.g. Chromatics MPACT) have emerged. These architectures

- 2 -

use identical processing elements executing the same instruction to perform the same processing on a number of blocks of data in parallel. This approach works well for data which can be easily partitioned to allow a common function to be performed e.g. block processing in data compression such as MPEG, but are not flexible enough to execute a complete general algorithm which often requires conditional flow control within the data processing.

DSP (digital signal processor) vendors have also sought to address this market with MIMD (Multiple Instruction Multiple Data) devices (e.g. Texas Instruments' TI320C80) which offer the required flexibility to process the varied data types. However since the architecture replicates general purpose DSP cores which retain a far greater degree of flexibility than required for the application, the resulting chip is a high cost device, too high for general PC and consumer use.

CPU (central processing unit) vendors promoting fast RISC CPUs for both general purpose programs and multimedia processing are unable (and do not wish) to compromise their architecture in order to support more than a few multimedia specific instructions and therefore do not achieve the required performance levels at a reasonable cost. As the CPU is also typically being used to run a non-real-time operating system, it is also unable to provide low latency processing.

- 3 -

Dedicated multimedia CPUs (e.g. Philips' Trimedia) using VLIW (very long instruction words) instructions controlling multiple processing units are unable to make efficient use of their processing power because each instruction is dedicated to a single task (and data type) and therefore unable to make optimal use of all the processing units available. For example a VLIW instruction dedicated to a 3D graphics operation is unable to take advantage of hardware designed for MPEG motion estimation. The number of processing units, and therefore scale-ability, is also limited by the VLIW word length.

SUMMARY OF THE INVENTION

Preferred embodiments of the present invention address the requirement for a device which processes all multimedia data types in a manner that minimises system costs and provides for future developments in multimedia and the related industry standards. They provide an architecture which is scalable in processing power, real-time I/O support and in the number of concurrent activities which can be undertaken.

All multimedia data types may be viewed as streams of data which lend themselves to a vector processing approach. Some of these streams will be real time (e.g. from an audio or video input) and as such either require dedicated buffering or low latency processing to avoid data loss. Each data

- 4 -

stream also requires some hardware resource so that it may be processed.

A preferred embodiment of the invention includes a low latency real-time processing core responsible for data IO and task scheduling only. This avoids the need for unnecessary and costly buffering. It also includes a method of dynamic resource checking to ensure that only tasks with the required resources available are run.

The balance between host processing power, memory costs and silicon costs is also continually changing. This means that the optimal division of work between a host processor and multimedia coprocessor also changes over time. This device is programmable to allow the division of work to be altered as required.

Scale-ability of parallel processing devices is a problem for both hardware design and supporting software. As more processing units are added to a device the distribution of tasks between the processing units becomes more difficult resulting in either a diminishing return or an exponential growth in the number of inter-connects between functional units. Such changes also typically result in alterations to the programming model for the device requiring wholesale changes to the supporting software. Preferred embodiments of the invention address these issues by a consistent scalable architecture, where all the elements may be scaled without creating an explosion of inter-connects

- 5 -

between functional units and without changing the programming model presented to software interfacing to the device.

Figure 1 shows the base architecture of the
5 device.

The device has been conceived as a re-configurable engine able to match all the current and future algorithms required to process multimedia data. The work done by it is split into two
10 categories. Both real time scheduling and IO processing are performed by a Media Control Core whilst computationally intensive data processing is performed by one or more additional data processing units.

15 This division of work is one of the architecture's fundamental characteristics.

Data processing consists of a number of steps:

Parameter fetching and setup
Data fetching and processing
20 Data storage

In order to efficiently achieve high data processing throughput a processor needs to perform the above operations on a reasonably large set of data. If the data set is too small the processor
25 spends too high a proportion of it's power on context switching between tasks and the resulting need to save and restore a thread's state.

Because the Media Control Core is required only to service requests to move data between IO ports and
30 memory (to allow data processing to be performed) it

- 6 -

can context switch every clock cycle, this then removes the need for large data buffers to support real time IO. Data processing units are able to process data efficiently by performing a key part of
5 an algorithm on data without interruption.

These processing elements are supported by a scalable multibank cache which supports efficient data movement and processing by caching sets of data required for the active algorithms being run.

10 The invention is defined in its various aspects with more precision in the appended claims to which reference should now be made.

A preferred embodiment of the invention will now be described in detail, by way of example, with
15 reference to the figures in which:

Figure 1 shows a block diagram of an embodiment of the invention;

Figure 2 shows a block diagram of the Media Control Core of Figure 1;

20 Figure 3 is a block diagram of a second embodiment of the invention;

Figure 4 is a block diagram of the control unit instruction pipeline the Media Control Core;

25 Figure 5 is a block diagram of the internal architecture of one of the data banks of Figure 4;

Figure 6 shows in block form how resource checking and thus process selection is performed by the Media Control Core; and

30 Figure 7 is a block diagram showing how access is made to the banked cache memory of Figure 1.

- 7 -

The base architecture of the embodiment of the invention is shown in Figure 1. The centre of the system is a media control core (MCC) 2. This is a fine grained multithreading processor. This has a plurality of inputs and outputs which can be coupled to real time data input and output devices 4. These can be, for example, video sources, audio sources, video outputs, audio outputs, data sources, storage devices etc. In a simple example only one input and one output would be provided.

Also coupled to the media control core 2 are a plurality of data processing units 6. Each of these comprises a data processing core 8 which controls the processing of data via data pipeline 10. The core 8 decodes and sequences microinstructions for the pipeline 10.

Also coupled to the media control core 2 is a multibanked cache memory 12 from which data may be retrieved by the media control core 2 and data processing units 6 and into which data may be written by the media control core, the data processing units 6.

The media control core is a fine grained multithreading processing unit which directs data from inputs to data processing cores or to storage and provides data to outputs. It is arranged so that it can switch tasks on every clock cycle. This is achieved by, on every clock cycle checking which of the possible operations it could perform have all the resources available for those tasks to be executed

- 8 -

and, of those, which has the highest priority. It could be arranged to commence operation of more than one operation on each clock cycle if sufficient processing power were provided.

5 This resource checking ensures that everything required to perform a particular task is in place. This includes external resources such as whether or not data is available at an input port (EG video data) or whether a data storage device or output is
10 available. It also includes internal resources such as data banks for temporary storage, available processing cores which are not currently working on other data or previously processed data required for a particular new processing operation. The media
15 control core operates to direct data from an input to an appropriate data processing unit 6 for processing to take place and routes data to an output when required making use of the cache as necessary. Once
20 execution of a set of instructions has commenced on a processing unit the MCC can look again at the various threads it can run and the resources available for these whilst the program continues to run on the data processing unit.

 The resource and priority checking of the media
25 control core means that tasks which serve as real time data such as video input are able to be performed without the large memory buffers which are usually required in current real time inputs. In
 operation such as video input the media control core
30 will look to see whether data is available at the IO

- 9 -

port and, if it is, will receive that data and send it either to a portion of the multibanked cache or to data storage registers in preparation for processing by the one of the data processing unit 6.

5 The data processing units 6 are all under the control and scheduling of the media control core 2. In the example shown in Figure 1 the units consist of a processing pipeline (data pipeline 10) which will be made up of a number of processing elements such as multipliers, adders, shifters etc under the control
10 of an associated data processing core 8 which runs a sequence of instructions to perform a data processing algorithm. Each of these data processing cores will have its own microinstruction ROM and/or RAM storing
15 sequences of instructions to perform a particular data processes. The media control core invokes the data processing unit 6 to perform its particular operation sequence by, for example, passing an address offset into its microinstruction ROM and
20 instructing it to commence execution. It will then perform a particular process on either data from the multibanked cache or data passed to it from one of the inputs to the media control core until completed when it will signal to the media control core that
25 its processing is complete.

 The multibanked cache 12 of Figure 1 is used for memory accesses and these are all cached through this bank. The cache is divided into a plurality of banks 14 each of which can be programmed to match the
30 requirements of one of the data processing tasks

- 10 -

being undertaken. For example, a cache bank might be dedicated to caching texture maps from main memory for use in 3D graphics rendering. Using this programmability of the cache banks allows the best possible use of on chip memory to be made and allows
5 dynamic cache allocation to be performed thereby achieving the best performance under any particular conditions.

Furthermore, the use of multiple cache banks
10 allows the cache to be non-blocking. That is to say, if one of the cache banks is dealing with a request which it is currently unable to satisfy, such as a read instruction where that data is not currently available, then another processing thread which uses
15 a separate cache bank may be run.

The entire device as shown in Figure 1 is scalable and may be constructed on a single piece of silicon as an integrated chip. The media control core 2 is scalable in a manner which will be
20 described below with reference to Figure 2. As the size of the media control core is increased it is able to support further data processing units 6 whilst using the same programming model for the media control. More cache banks may also be added to
25 support the further data processing units thereby increasing the effectiveness of the data throughput to the media control core and the data processing units. Because the programming model of the device is not changed this enables a high degree of
30 backwards compatibility to be attained.

- 11 -

The media control core is shown in more detail with reference to Figure 2. It is composed of a control unit 16, a set of read/write units 18, a set of program counter banks 20, a set of address banks 22, a set of data banks 24, and a set of input/output banks 26. These banks are all coupled together by a media control core status bus 28 a media control core control bus 29 and a media control core data interconnect 30. The media control core data interconnect is used for sending data between the various different banks and the status bus provides data such as the input/output port status and the status of data processing units to which the media control core can send instructions and data.

In addition, a memory block 32 storing microcode instructions in ROM and RAM is coupled to the control unit 16 the units 18 to 26 listed above.

All the core components, 18 to 26, with the exception of the control unit have the same basic interface model which allows data to be read from them, written to them and operations performed between data stored in them. Each bank consists of a closely coupled local storage register file with a processing unit or arithmetic logic (ALU).

The control unit 16 is used to control the execution of the media control core. On each clock cycle it checks the availability of all resources (e.g. input/output port status, data processing units status, etc) using status information provided by the media control status bus 28 against the resources

- 12 -

required to run each program under its control. It then starts execution of the instruction for the highest priority program thread which has all its resources available.

5 The program counter bank is used to store program counters for each processing thread which is supported by the media control core. It consists of a register for each of the processing threads which the media control core is capable of supporting and
10 an ALU which performs all operations upon the program counters for program progression, looping, branching, etc. The data banks 24 are used for general purpose operations on data to control program flow within the media control core. They are a general resource
15 which can be used as required by any processing thread which is running on the MCC.

 The address banks 22 are used to store and manipulate addresses for both instructions and data and are also a general MCC resource in a similar
20 manner to the data banks 24.

 The input/output banks 26 provide an interface between the media control core and real time data streams for input/output which are supported by the MCC. Their status indicates the availability of data
25 at a port, eg. video input, or the ability of a port to take the data for output. They can, as an option, include the ability to transform data as it is transferred in or out, for example bit stuffing of a data stream.

- 13 -

The read/write banks 18 provide an interface between the media control core and memory (via the multibank cache). As more than one processing thread can be run at any one time more than one read/write unit is required to avoid the blocking of memory requests.

The media control core is scalable in all important respects. Because it is constructed from banks which localise storage (register files) and processing (ALU) additional banks can be added without creating any unmanageable routing and interconnection problems. The number of processing threads which could be supported can be increased by adding registers to the program counter bank and modifying the control unit accordingly. The number of input/output streams which can be supported by the MCC can be increased by adding further IO banks.

The data throughput can be increased by adding further read/write units 18 and the MCC processing power overall can be increased by adding further data and address banks, 24 22.

A block diagram of a specific implementation of the data processing management system is shown in Figure 3. The MCC in this serves as a plurality of real time data input/output ports and controls data processing units to process data received from them and output to them.

In the figure is shown a video input 34 and audio input 36 coupled to the media control core via associated preprocessors 38 and 40. A corresponding

- 14 -

video output 42 and audio output 44 are coupled to the media control core 2 via respective post processors 46 and 48. The video and audio inputs and outputs may be digital inputs and outputs.

5 As in Figure 1 the media control core 2 is coupled to a multibanked cache 12 in this case referred to as the main cache bank. A data processing unit 6 comprising a secondary core 8 and a data (media) pipeline 10 are coupled directly the
10 media control core and are used for processing of data supplied to them.

Also coupled to the media core 2 is a processing unit 50 comprising a digital to analog converter feed core (DAC feed core) 52 and a DAC feed pipeline 54
15 which supplies data to a digital to analog converter 56. The purpose of this is to provide a graphics output. To this end, the processing unit 50 fetches data via the frame buffer interface 58 and system bus 60 for the host computer video graphics adaptor (VGA
20 62) is retained for compatibility only. Thus, real time data is supplied on the video and audio inputs and can be sent out on the video and audio outputs whilst graphics output can be sent by the DAC 56.

Data for graphics output can be generated by
25 processing non-real time data from a source such a graphics frame buffer, a connection to which is shown in Figure 3 via the frame buffer interface 58, 3D data, or real time video.

The secondary core 8 and media pipeline 10 is an
30 example of a data processing unit which is able to

- 15 -

process audio, 3D, 2D, video scaling, video decoding etc. This could be formed from any type of general processor.

The DAC feed core and DAC feed pipeline is
5 dedicated to processing data from a number of frame buffers for the generation of RGB data for a DAC. It can switch between source buffers on a pixel by pixel basis, thus converting data taken from a number of video formats including YUV and combining source data
10 from multiple frame buffers by blending or by colour or chroma keying.

Each core will have an associated microcode store formed from ROM and RAM which for the purposes of clarity are not shown here, but which stores
15 instructions to be executed by the processor. The cache banks 12 interface to the media control core and the data processing units 6 and 50. They also interface to the system bus via an address translation unit 64. They are also linked to the
20 frame buffer interface 58 for writing data to an reading data from one or more frame buffers.

A data bank 24 is illustrated in Figure 5. It comprises a register file 72, an ALU 74, and a multiplexed input 76. The operation of the data bank
25 is controlled by a number of bits in a micro-instruction which are labelled WE, W, R1, and R2 and which are input to the register file. The result of the micro-instruction which is performed by the ALU is made available as status bits H S Z which are
30 routed to the control unit of the media control core

- 16 -

to implement branches and conditional instructions.

The register file is constructed to allow two operands to be fetched from the input and one operand to be written to the output on each clock cycle. The data input port 78 and the data output port 80 allow communication with other data via the media control core data bus 30 to which they are connected. Thus, the data flow in Figure 5 is vertically down through the diagram whilst the flow of control information is from left to right being formed of control bits from the control unit and status bits sent back to the control unit reflecting the status of the data bank.

A plurality of these data banks are used and each is in the same form, that is to say each has its own register file closely coupled to an ALU as shown in Figure 5. This arrangement, using a plurality of closely coupled registers and ALU's, preferably in a one to one relationship, differs from prior art embodiments of multiple ALU's where complex multiplexing between register banks and multiple ALU's was required.

Generally, these data banks perform general purpose operations on data thereby controlling program flow within the MCC and can be used by any processing thread which is running on the MCC.

The address banks 22, the program counter banks 20, and the IO banks 26, and the read/write units 18 are all constructed and operate in a similar manner but are provided in separate units to allow their

- 17 -

implementation to be optimised, thereby reflecting the way in which they are used.

The address banks store and manipulate addresses for data accesses into memory (not illustrated).

5 They are slightly simpler than the data banks in that they use unsigned accumulators and do not generate any condition codes to send back to the control unit 16 via the status bus.

The program counter bank is used to store the
10 program counter for each processing thread supported by the média control core. Thus, the number of registers in the bank of the type shown in Figure 5 will be equivalent to the number of processing threads which the MCC can support. As with the
15 address banks the ALU is used to program counter operations and is unsigned. It does not generate conditions codes to send back to the control unit 2.

The IO banks are used to interface to IO ports, and contain no registers or ALU's. They interface
20 with real time data streams supported by the MCC. A status signal indicates the availability of data at a port, or the ability of a port to take data. They can optionally include the ability to transform the data as it is transferred.

25 The read/write units interface to the cache bank 12. They have no registers or ALU's. A read unit accepts an address and, when the data is returned, sets a data valid status bit. A write unit accepts addresses and data. Multiple read and write units
30 are used to ensure that if one cache access blocks

- 18 -

then another thread can be continued running through another read/write unit.

An instruction buffer with the control unit (not illustrated) for each data processing thread stores that thread's next microinstruction and instruction
5 operands. The instruction and operands include bits which describe the resources required to execute that instruction. These resource requirements are fed into the control unit's resource checking logic along with
10 status bits describing the current status of the Media Control Core 2, external IO ports 20 and data processing units 6,50. Simple combinatorial logic such as an array of logic gates determines whether an instruction can run or not and a fixed priority
15 selector in the control unit 16 then launches the highest priority runnable thread into the data path control pipeline (shown in Figure 4) to start execution of that program thread. The threads task could be 'receive video data', process stored audio
20 data' etc.

Normally an instruction will request its thread's next instruction to be read from memory when it is run. The instruction is read from memory (pointed to by the program counter) which contains an
25 instruction opcode and operands. The opcode field of the instruction is used to index into the microcode ROM to retrieve the next instruction and the resultant microinstruction is stored into the thread's instruction buffer together with the
30 instruction operand fields.

- 19 -

The resource checking and priority is illustrated fully in Figure 6. For the three threads illustrated, global status information is received from the necessary data banks, the necessary address banks, routing control data from the control unit, control status information from control unit 16, and execution dependency data from other processes on which a particular thread is dependent. All this information is sent to a resource checker 80 which combines it with data from IO ports, the various pipeline data bank status, and the status of the various data processing units. This happens for each possible thread. If it is possible to run that data processing thread then an output is generated to a priority selector 82. This has information about the priority of each of the data processing threads supported and, as a result, can select for execution the thread with highest priority. For example, a real time data input such a video would be given a high priority and this would take precedence over a background processing operation.

Because the next instruction for a thread is already provided in an instruction buffer that instruction is always available for resource checking and priority selection. Thus, there is no loss of execution time by checking the status of every clock cycle.

The data path control pipeline shown in Figure 4 operates by allowing fields of a microinstruction word to be placed into a pipeline at different

- 20 -

depths. This allows a microinstruction to control the flow of data through the pipeline over a number of clocks and hence to control the pipelined processing of data.

5 The circuitry of Figure 4 comprises a 1 to 4 decoder 90 which on its enable input receives the output of an AND gate 92. The inputs to this are a control bit from the microcode instruction and a condition code used for conditional execution of
10 instructions. A pair of timing bits from the microcode instruction which are the output selection inputs to the decoder 90. The four outputs of the decoder 90 form inputs via OR gates 94 to four D-type flip-flops 96 arranged as a shift register. The
15 outputs from decoder 90 are ORed in gates 94 with the outputs of the previous flip-flop 96 in the register (output from the first flip-flop 96). Bits are clocked along the register by a clock 98 until they emerge as an output control bit which commences
20 execution of the microcode instruction.

 Thus a control bit is inserted into the correct position in its scheduling pipeline such that it arrives at the destination bank on the required clock cycle. Conceptually such an instruction bit pipeline
25 exists for all microcode control bits but in order to limit the amount of logic needed to implement the control pipeline, there are limitations on the clock cycles on which some fields of the microcode can be placed.

- 21 -

Conditional execution is achieved by specifying a conditional operation and generating the condition bit. Two types of conditional operation are supported. This first is to qualify the write enable pulse to a bank with the condition code from the same or another bank. The second is to specify that a microinstruction word would be run again (rather than the next instruction from the program counter) if a certain condition code is true. In order to limit the number of possibilities for condition codes, only data bank condition codes can be used in these conditional operations.

Example Microinstruction Format

The following gives an example of a microinstruction format for this architecture and explains how it is used to achieve multithreading on a cycle by cycle basis.

In Figure 6 a number of thread's microinstructions are shown. Each contains the following:

Control fields for each bank e.g. Register select bits and ALU control bits;

Instruction timing bits for each bank - these are explained below;

Routing control bits which control routing of data between banks;

Core control bits such as whether the instruction should be conditionally repeated and whether it contains immediate data operands.

- 22 -

For performance, instructions are allowed to execute over a number of clock cycles. The time at which parts of the instruction executes is set by delay bits within the bank control field which control the position that the control bits are placed in the Control Unit Instruction Pipeline (Figure 6).

Because the control bits have been placed in the Instruction Pipeline which represents the future state of the Media Control Core, the control unit logic is able to ensure that the instruction delay is catered for when resource checking and that an instruction will cause no conflicts on any of the clock cycles in which it is executing.

Execution Dependencies

In order to keep the hardware design complexity down, instructions are allowed to execute over a number of clock cycles. The time at which parts of the instruction executes is controlled by delay bits within the bank control field.

In order to ensure both that this instruction delay is catered for each clock when the resource checking is undertaken and that the operation happens on the correct cycle the op-code corresponding to the delay bits is fed into a set of latches which are clocked each cycle. The outputs of these latches represent the future state of the data pipeline and are fed into the resource checking logic to ensure that an instruction will cause no conflicts on any of the clock cycles in which it is executing.

- 23 -

Banked Cache

The multibanked cache is formed from a number of cache banks and interfaces to processing units and memory as shown in Figure 7. In order to support an arbitrarily scalable device a multiplicity of cache banks are used. The use of each bank is controlled by a cache allocator 86 associated with a cache user such a read unit, or a write unit. These may be programmably controlled to use the cache banks in different configurations. For example, one bank may be used for command data, another for 3D texture maps, and a third for 2D parameters. The ability to configure the cache banks is important in achieving good memory performance.

Each port such as a read unit 88 or a write unit 90 which requires access to the cache is connected to an allocator module 86. These modules examine the memory request that is being made by the port and route the request to the appropriate cache bank. The address sent from the port is compared with a base range register pair in the write allocator to determine whether or not the address falls within a given region. If a match occurs then the request is forwarded to the cache bank. If no match occurs, a default cache bank is used. This comprises simply passing the request through to the memory sub-system.

More than one set of base and range registers may be used, depending on the memory requirements of the module connected to the port.

- 24 -

Not all of the cache banks provided need to be accessible from every given allocator. Some ports will need more flexibility than others. This fact allows the number of cache banks to be easily scaled
5 (increased) whilst restricting the growth of interconnections required between allocators and cache banks. Thus, a set of caches might be allocated to deal with video input requests and audio inputs and outputs whilst others could be allocated
10 to deal primarily with data fetches from main memory.

Each cache bank is connected to read and write allocators via an arbiter 94. This receives requests for access from all of the allocators and can then determine which allocator is to obtain access to that
15 particular cache bank. This is done by assigning a priority to each port and arranging for the arbiter to simply process the highest priority request that is outstanding.

The system can be extended to use other types of
20 inputs such as MPEG and video conferencing.

- 25 -

CLAIMS

1. A data processing management system comprising:

at least one data input;
at least one data output;
5 at least one data processing means;
a data storage means; and,
a control means;

wherein the control means comprises:

means for routing data between the data
10 input, the data output, the data processing means and
the data storage means in one or more programmable
routing operations;

means for causing the data processing means
to commence a predetermined data processing
15 operation;

means for repeatedly determining which
routing operations and which data processing
operations are capable of being performed;

means for commencing execution of at least
20 one at the thus determined operations capable of
being performed.

2. A data processing management system according to
claim 1 including means for assigning a priority to
each routing and data processing operation; and

25 means for determining which of the routing and
data processing operations capable of being performed
has the highest assigned priority, wherein the means
for commencing execution is controlled to commence

- 26 -

execution of the operation with the highest assigned priority.

3. A data processing management system according to claim 1 or 2 in which the data input is a real time
5 input and the operation of receiving data on that input is assigned the highest priority.

4. A data processing management system according to claim 3 in which the data input is a video data input.

10 5. A data processing management system according to claim 3 in which the data input is an audio input.

6. A data processing management system according to any of claims 2 to 5 in which the means for repeatedly determining which routing operations and
15 which data processing operations are capable of being performed and the means for determining which of the thus determined operations has the highest assigned priority make this determination on each clock cycle of a clocking means associated with the control
20 means.

7. A data processing management system according to claim 6 in which the commencement of the thus
determined operation with the highest assigned
priority takes place on a succeeding clock cycle.

- 27 -

8. A data processing management system according to any preceding claim in which the means for determining which routing and data processing operations are capable of being performed makes this
5 determination from resource status bits received via a status bus and generated by internal and/or external resources.

9. A data processing management system according to any preceding claim in which the data processing
10 means includes a store of microcoded instructions relating to a processing operation to be performed by the processing means.

10. A data processing management system according to claim 9 in which the control means commences
15 execution of the processing operation by providing an address offset into the microcode instruction store of the data processing means.

11. A data processing management system according to any preceding claim in which the data storage means
20 comprises a cache memory means.

12. A data processing management system according to claim 11 in which the cache memory means comprises a
plurality of banks of cache memory storage.

13. A data processing management system according to
25 claim 1 in which each portion of the system which has

- 28 -

has access to the cache memory is associated with a cache memory allocation means which is programmable to permit access to different banks of cache memory.

14. The data processing management system according to any preceding claim in which the control means comprises a set of data banks for performing operations on data within the control means.

15. A data processing management system according to any preceding claim in which the control means comprises a set of address banks, one for each of the data processing operations to be performed by the system.

16. A data processing management system according to any preceding claim in which the control means comprises a program counter bank for storing the current program address for each data processing operation to be performed by the system.

17. A data processing management system according to any preceding claim in which the control means comprises a set of input/output banks for interfacing with the input and output means.

18. A data processing management system according to any preceding claim in which the control means comprise read/write units for interfacing with the storage means.

- 29 -

19. A data processing management system according to claim 12 in which each data bank comprises an arithmetic logic unit (ALU) and a register file associated only with that ALU.

5 20. A data processing management system according to claims 14 to 19 in which the data banks, the address banks, the program counter banks, the input/output banks, and the read/write banks are all connected to a common status bus a common data interconnect and a
10 common control bus.

1 / 5

FIG. 1

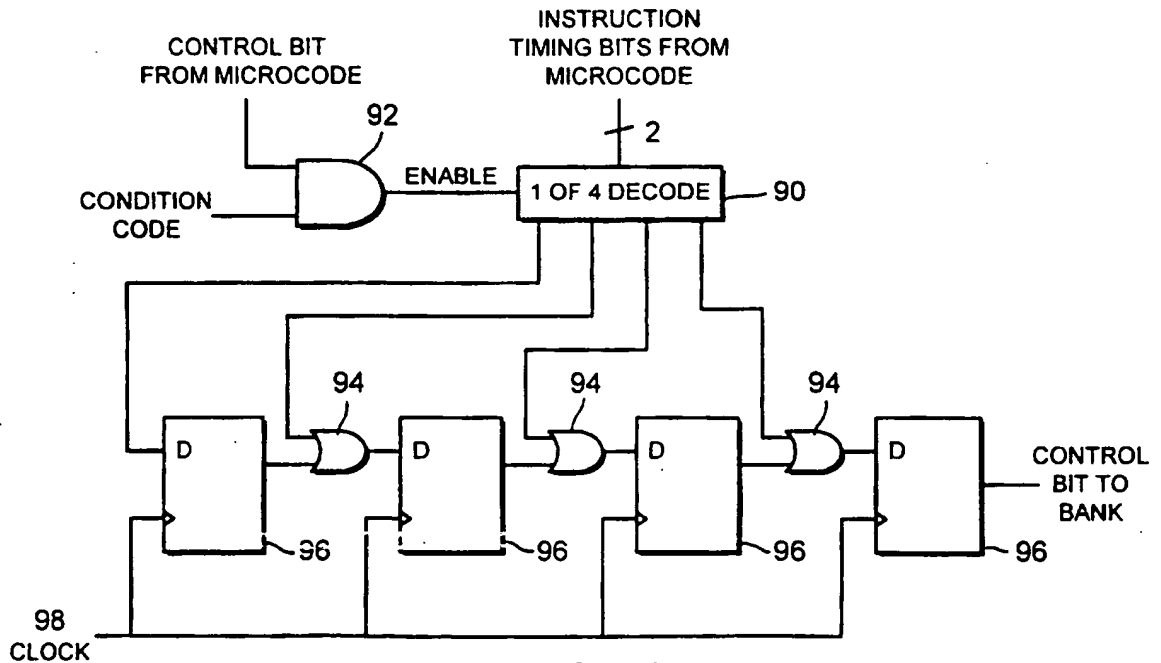
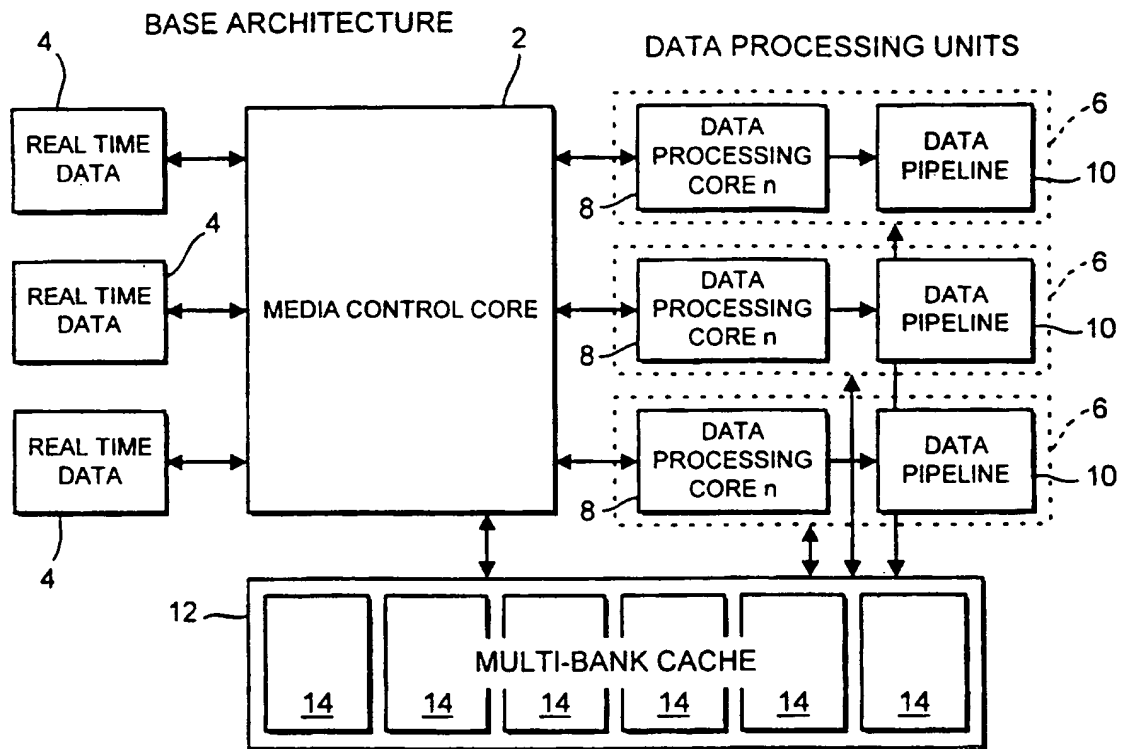


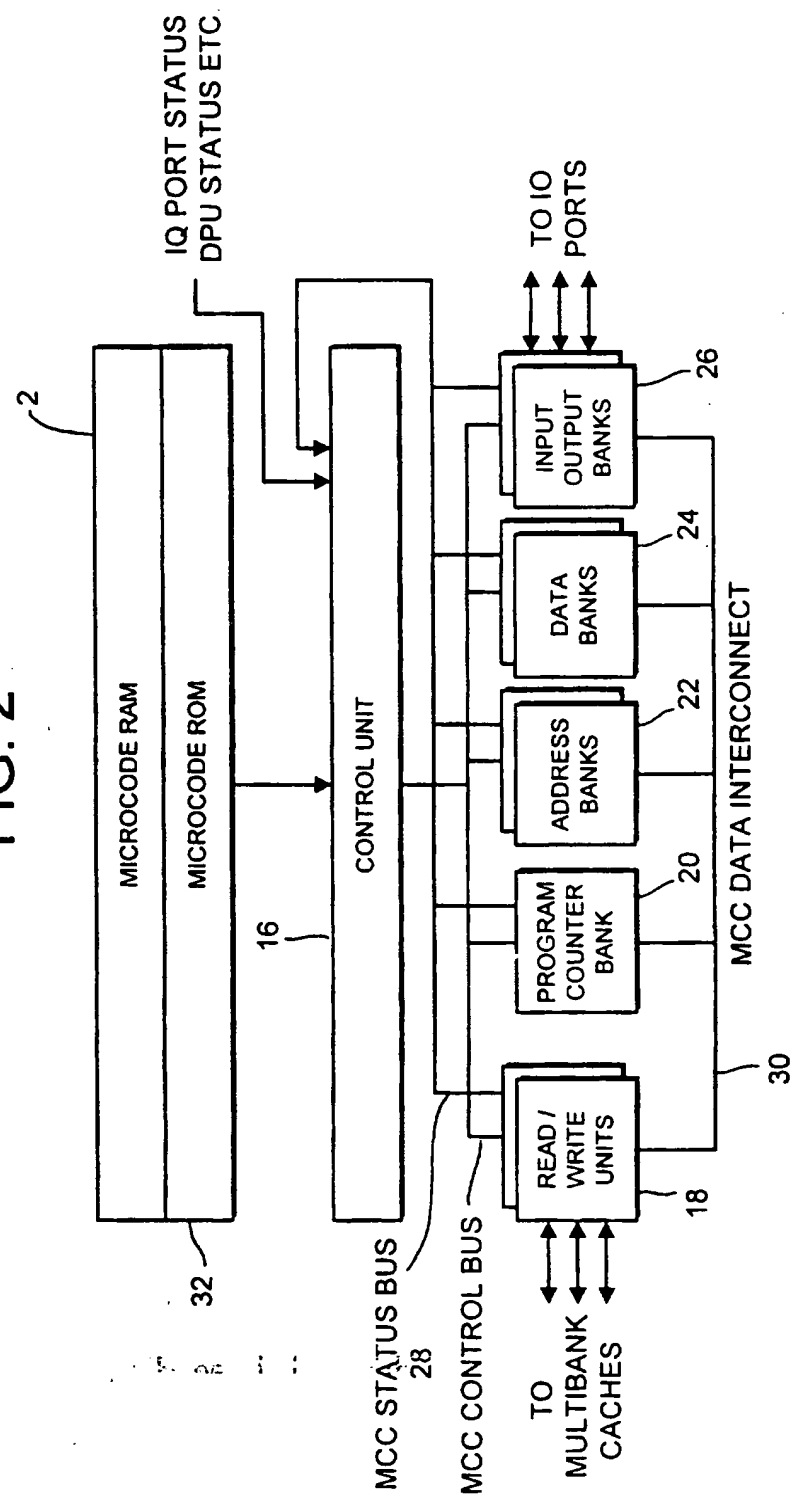
FIG. 4

CONTROL UNIT INSTRUCTION PIPELINE

SUBSTITUTE SHEET (RULE 26)

2 / 5

FIG. 2



3 / 5

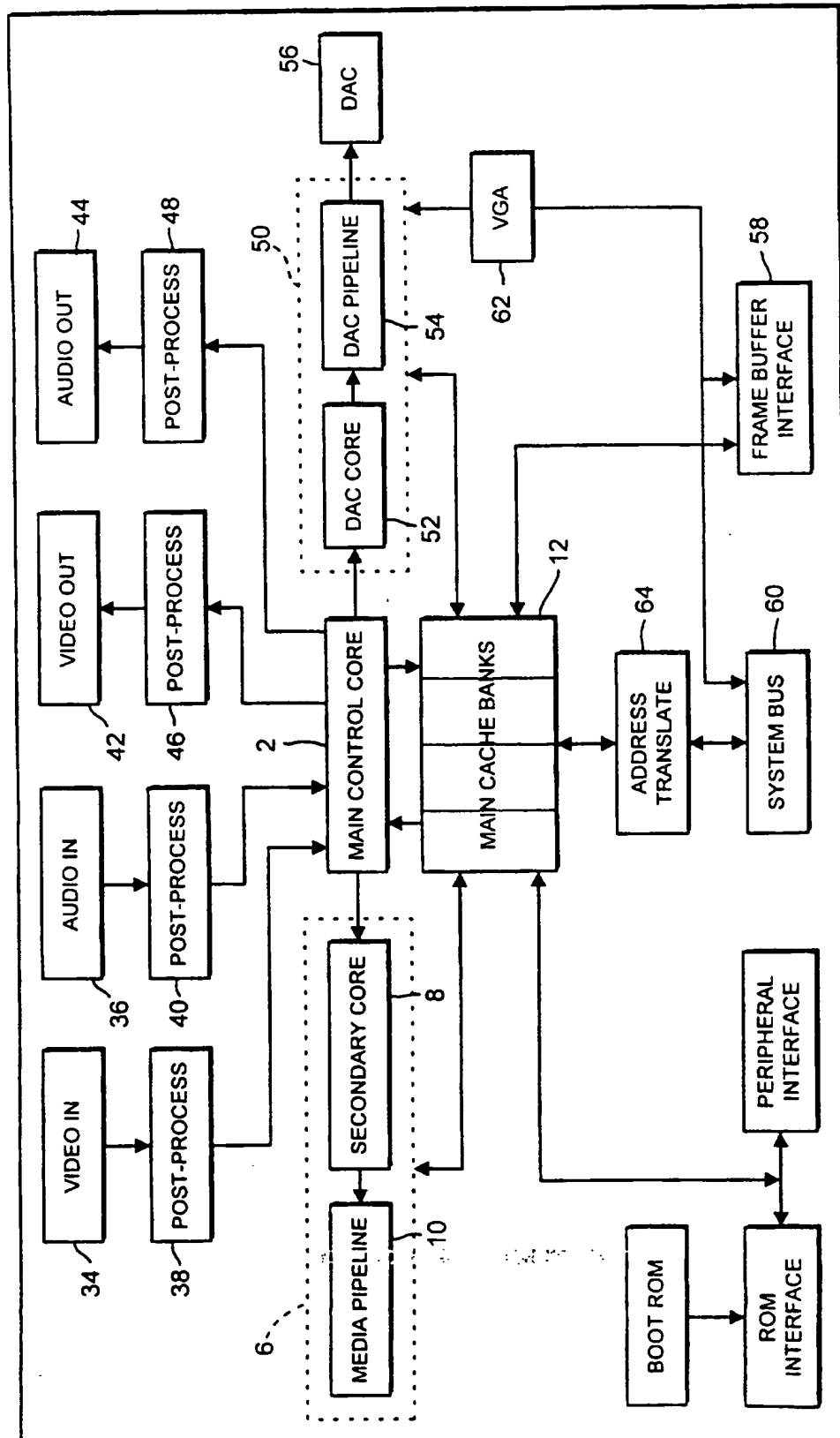


FIG. 3
SCALEABLE MULTIMEDIA PROCESSOR BLOCK DIAGRAM

4 / 5

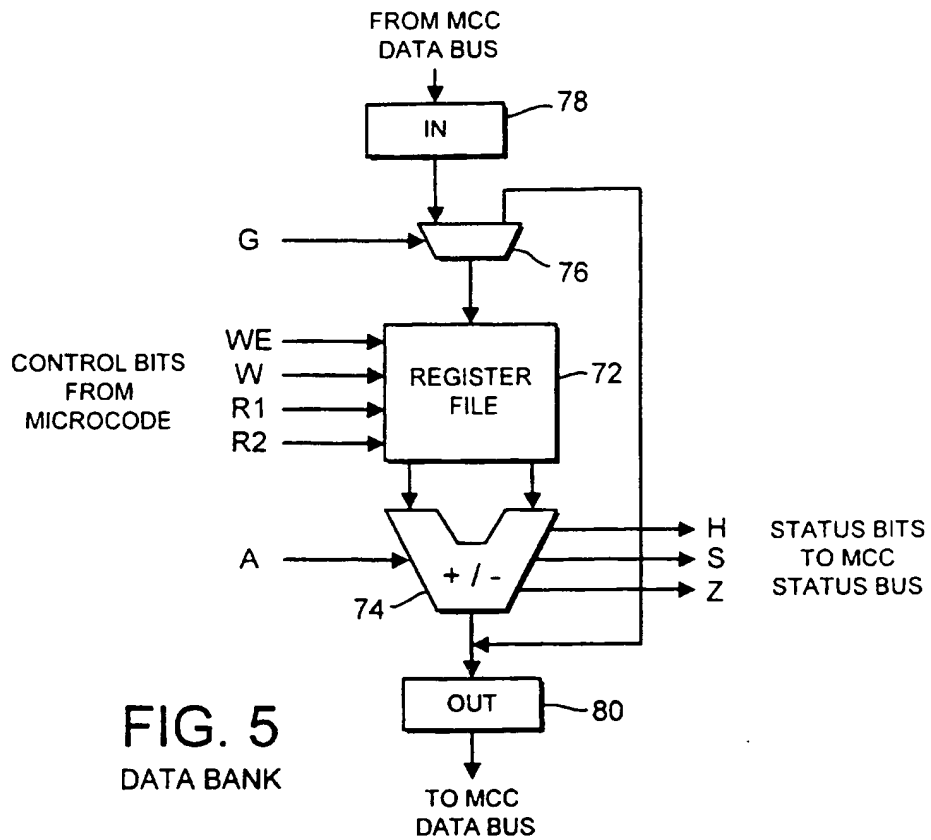


FIG. 5
DATA BANK

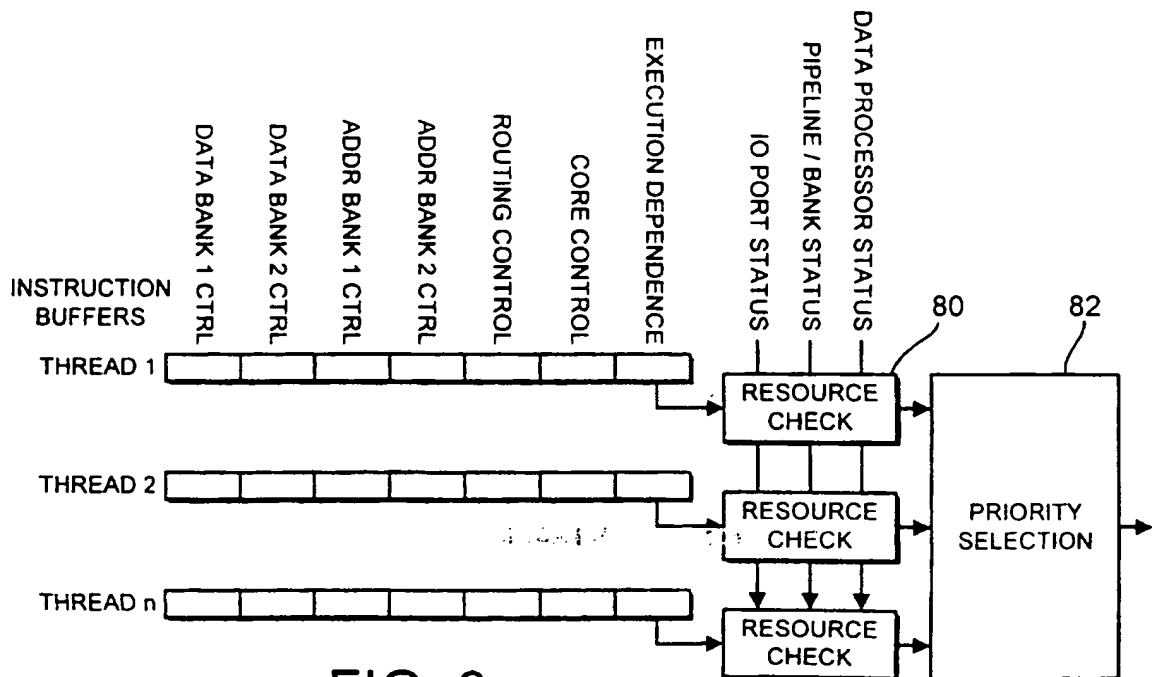


FIG. 6
RESOURCE CHECKING AND PRIORITY
SUBSTITUTE SHEET (RULE 26)

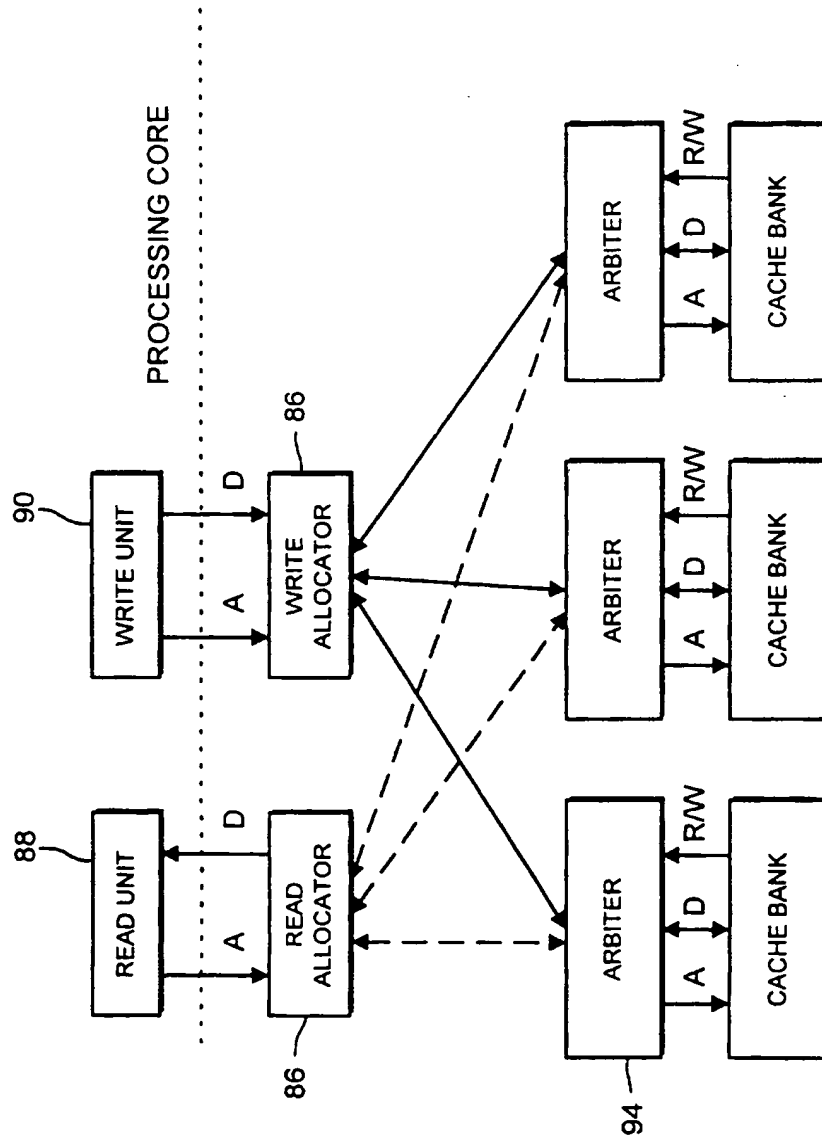


FIG. 7
BANKED CACHE ARCHITECTURE

INTERNATIONAL SEARCH REPORT

Inter: nal Application No
PCT/GB 97/00972

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 G06F15/80

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category * | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|------------|---|-----------------------|
| X | US 5 487 153 A (HAMMERSTROM DANIEL W ET AL) 23 January 1996 see column 1, line 30 - column 2, line 22; figure 1 see column 11, line 57 - column 12, line 52 --- | 1 |
| A | EP 0 020 202 A (THOMSON CSF) 10 December 1980 see page 3, line 2 - line 29 see page 7, line 35 - page 9, line 20 --- | 1,8 |
| A | EP 0 367 639 A (EVANS & SUTHERLAND COMPUTER CO) 9 May 1990 see column 2, line 40 - column 5, line 20 see column 6, line 30 - column 8, line 8 --- | 1,2,7,8 |
| | -/-- | |

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

* Special categories of cited documents :

- * "A" document defining the general state of the art which is not considered to be of particular relevance
- * "E" earlier document but published on or after the international filing date
- * "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- * "O" document referring to an oral disclosure, use, exhibition or other means
- * "P" document published prior to the international filing date but later than the priority date claimed

- * "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- * "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- * "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
- * "&" document member of the same patent family

Date of the actual completion of the international search

22 July 1997

Date of mailing of the international search report

11.08.97

Name and mailing address of the ISA
European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+ 31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+ 31-70) 340-3016

Authorized officer

Michel, T

INTERNATIONAL SEARCH REPORT

International Application No
PCT/GB 97/00972

| C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT | | |
|--|--|-----------------------|
| Category | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
| A | PROCEEDINGS OF THE SUPERCOMPUTING CONFERENCE, ORLANDO, NOV. 14 - 18, 1988, no. 1988, 14 November 1988, INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS, pages 35-41, XP000042422 THISTLE M R ET AL: "A PROCESSOR ARCHITECTURE FOR HORIZON" see page 36, right-hand column, line 54 - page 37, left-hand column, line 34 --- | 2,6 |
| A | EP 0 397 180 A (MITSUBISHI ELECTRIC CORP) 14 November 1990 see the whole document --- | 1,4,5 |
| A | WO 94 15287 A (CENTRE ELECTRON HORLOGER ;PEROTTO JEAN FELIX (CH); LAMOTHE CHRISTI) 7 July 1994 see page 3, line 7 - page 4, line 34 ----- | 1,16 |

INTERNATIONAL SEARCH REPORT

information on patent family members

International Application No

PCT/GB 97/00972

| Patent document cited in search report | Publication date | Patent family member(s) | Publication date |
|---|---------------------|--|--|
| US 5487153 A | 23-01-96 | NONE | |
| EP 0020202 A | 10-12-80 | FR 2457521 A AU 538582 B AU 5866680 A CA 1150846 A US 4542455 A | 19-12-80 23-08-84 27-11-80 26-07-83 17-09-85 |
| EP 0367639 A | 09-05-90 | US 4985831 A CA 1323438 A JP 2178756 A | 15-01-91 19-10-93 11-07-90 |
| EP 0397180 A | 14-11-90 | CA 2016348 A JP 3205985 A US 5481727 A US 5237686 A | 10-11-90 09-09-91 02-01-96 17-08-93 |
| WO 9415287 A | 07-07-94 | CA 2128393 A CN 1089740 A EP 0627100 A JP 7504058 T US 5630130 A | 07-07-94 20-07-94 07-12-94 27-04-95 13-05-97 |